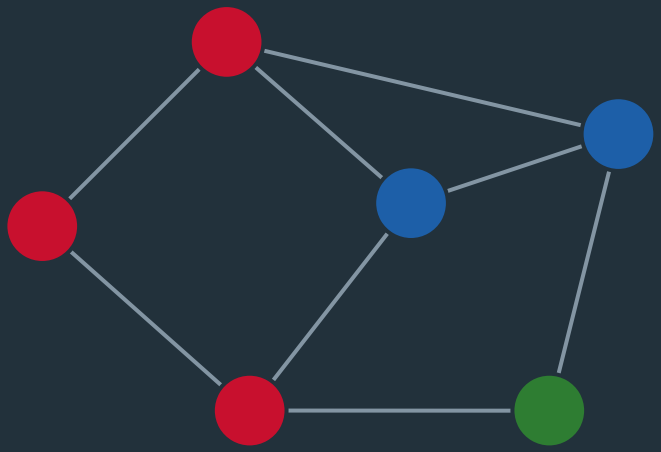
Ego networks

Course A · Week 9

Chapter 11 of Network Analysis with Stata (Stata Press) · one lecture and one lab, about 90 minutes each

Instructors are free to use, adapt and share these slides.



Today: a lecture, a break and a lab

0:00	**1 The circle as a unit of analysis**	20 min
0:20	**2 Size and density, and why the structure measures overlap**	30 min
0:50	**3 Composition, exposure and what they do not show**	40 min
1:30	Break	10 min
1:40	**4 Lab: size, density and circles**	45 min
2:25	**5 Lab: composition, exposure and promotion**	45 min

Two lawyers ask the same number of colleagues and live in different worlds

A closed circle

The advisers all ask one another. News reaches the lawyer from several of them, and each can check what another says.

An open circle

The advisers have never met. The lawyer sits between circles that would otherwise be apart.

A narrow or a wide circle

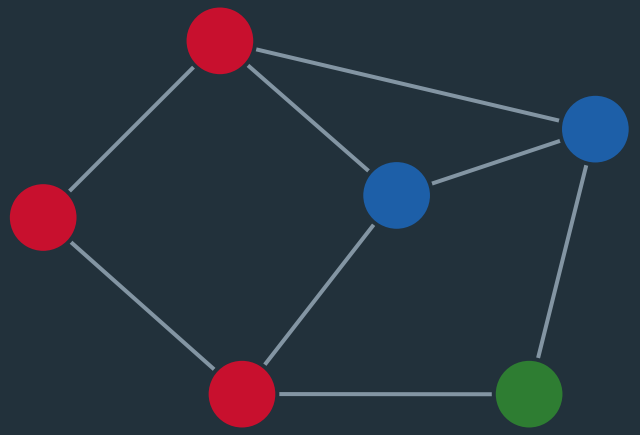
Only partners down the hall, or lawyers in all three offices: who the alters are is a separate question.

Same size, different circle. An ego network is one focal node, its alters, and the ties among the alters. Size alone describes only the first of those.

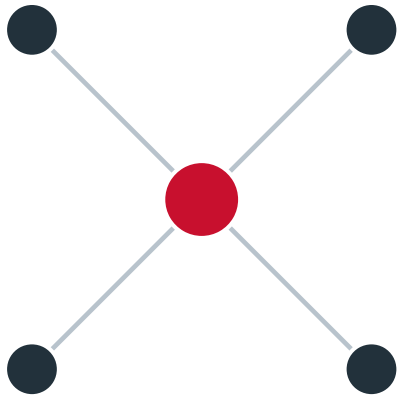
1

The circle as a unit of analysis

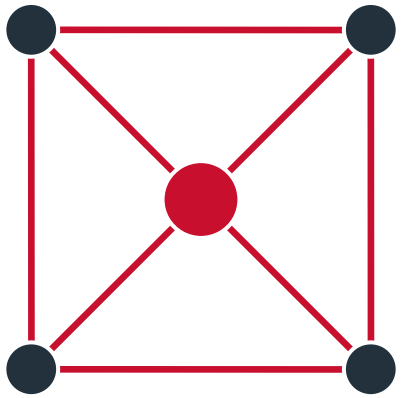
20 min · Chapter 11: ego, alters, the four things to measure, two ways to observe a circle



An ego network: one node, its alters, and the ties among them



Density 0: no alter knows another



Density 1: all six pairs tied

Red dot: ego. Dark dots: alters. Light lines: ego's ties. Red lines: ties among the alters.

Term	Meaning
Ego	the focal node
Alters	the nodes tied to ego
Size	the number of alters
Density	the share of possible ties among the alters that are present

Four things make an ego network a subject of its own

Structure

How many alters, and do they know each other? Size and density.

`nwego`

Composition

What are the alters like, and how far do they resemble ego?

`nwaltergen`

Exposure

The share of alters who do something: what ego may pick up from the circle.

`nwaltergen, hop()`

Design

Cut out of a whole network, or collected one respondent at a time.

`collapse`

The three arguments of Chapter 1, at the scale of one person

Argument	What it asks of the circle	Measure
Access to resources (Lin, Wellman)	many alters, of the right kinds	size, composition
Closure (Coleman)	alters who know one another: trust, mutual checking	high density
Structural holes (Burt)	alters who are not tied: ego bridges them	low density, effective size, constraint

Both readings need care. Density reflects the setting as much as the person: everyone in a small team knows everyone. And forty colleagues from one department make a large circle that gives access to little that is new.

Two ways to observe an ego network

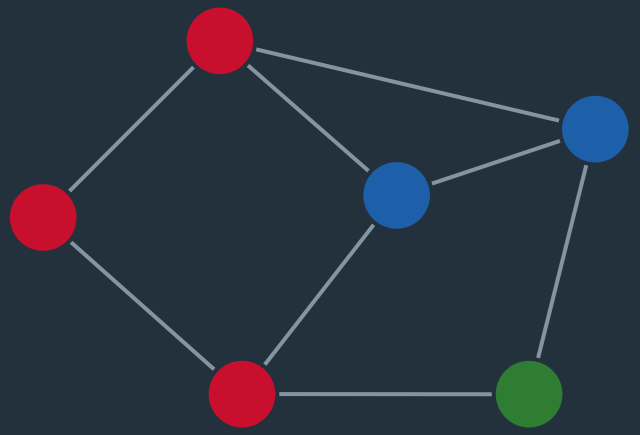
	Cut from a whole network	Personal-network survey
Alters	all contacts inside the boundary	the contacts the name generator elicited
Ties among alters	observed	reported by ego, with ego's perceptions and errors
Respondents	nodes: not independent observations	a probability sample: independent circles
Cannot be measured	nothing beyond what was observed	anything beyond the alters: two-step exposure, betweenness, communities

The measures do not care which design supplied them. The interpretation does: a name generator that asks about one kind of tie describes one part of the circle.

2

Size and density, and why the structure measures overlap

30 min · Chapter 11: nwego, nwegoplot, effective size and constraint



Size counts the alters; density asks whether they know each other

Size

The number of alters: a rough count of the people ego can draw on directly.

Density

The share of possible ties among the alters that are present. Missing for an ego with fewer than two alters: no pair to ask about.

Alters	Possible pairs
3	3
5	10
6	15
10	45

Read density next to size. Pairs grow with the square of the number of alters, so large circles are usually less dense.

nwego: 19.7 advisers on average, at a density of .315

```
. nwwebuse lazega, nwclear
. quietly nwdegree lazega_adv, generate(adv_out adv_in)
. nwego lazega_adv, sizevar(egosize) densvar(egodens)
    Variable |        Obs        Mean    Std. dev.       Min        Max
     egosize |         71    19.71831    8.634123          6         44
     egodens |         71    .3153674    .0845216   .1428571       .625
```

19.7

advisers on average (range 6 to 44)

- **Alters: anyone tied in either direction**

  The size is not out-degree plus in-degree: a reciprocated tie would count twice. The density is the share of ordered pairs of alters that are tied. Every lawyer has at least two alters, so density is defined for all 71.

.315

density of the average circle (range .143 to .625)

Size is almost out-degree; partners' circles are larger

```
. correlate egosize adv_out adv_in
    egosize |   1.0000
    adv_out |   0.9085   1.0000
     adv_in |   0.8083   0.5439   1.0000
```

.91

correlation of size with out-degree

.81

correlation with in-degree

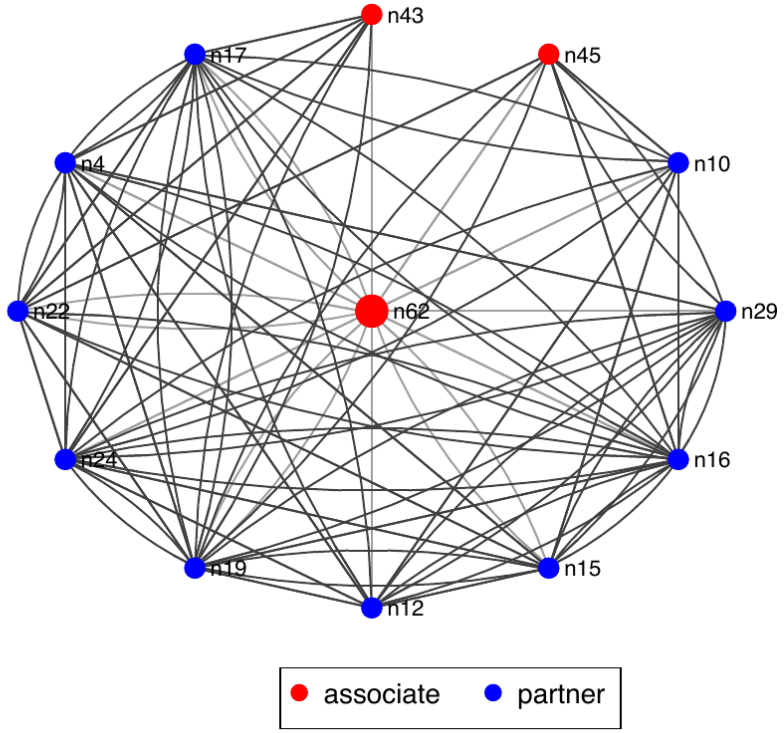
```
. tabstat egosize egodens, by(partner) statistics(mean sd n)
```

	mean size	mean density	lawyers
associates	14.4	.335	35
partners	24.9	.297	36
all	19.7	.315	71

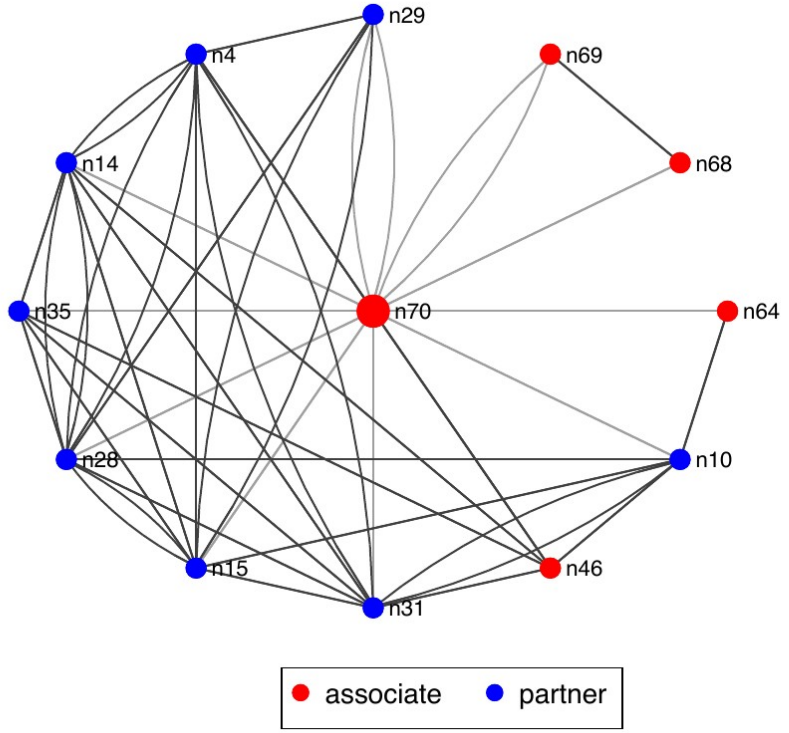
Rank goes with reach

Partners exchange advice with 24.9 colleagues, associates with 14.4. Larger circles are less closed: .297 against .335.

Two circles of 12 advisers: density .53 and .26



(a) n62: density .53, almost every alter tied to several others



(b) n70: density .26, three associates reach the partners via n70

```
. nwegoplot lazega_adv, ego(n62) label(_nwnode) color(partner) arrowfactor(0) scheme(s1network)
. nwegoplot lazega_adv, ego(n70) label(_nwnode) color(partner) arrowfactor(0) scheme(s1network)
```

Effective size and constraint largely repeat size and density

```
. nwwebuse florentine, nwclear
. nwego flomarriage, sizevar(fsize) densvar(fdens)
. quietly nwburt flomarriage
. list _nwnode fsize fdens _effsize if _nwnode == "medici", noobs
  | _nwnode   fsize      fdens   _effsize |
  |  medici       6   .0666667   5.666667 |
```

Undirected, ties present or absent: effective size = n - d(n - 1). The Medici have six partners with one tie among 15 pairs.

Correlations on lazega_adv	r
size, effective size	.90
size, constraint	-.69
density, constraint	.51
effective size, out-degree	.99

Correlations: run of 17 Nov 2026 (not printed in the chapter).

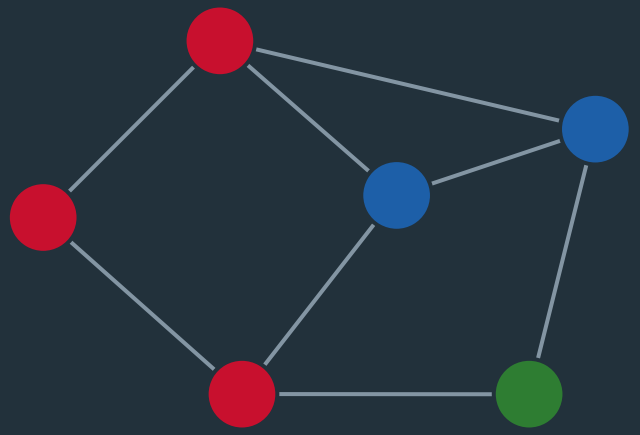
One summary in a regression

Size, density, effective size and constraint come from the same alters and ties. Enter one summary of structure and compare models.

3

Composition, exposure and what they do not show

40 min · Chapter 11: nwaltergen, a baseline for composition, collapse on survey data, exposure as a covariate



nwaltergen turns the alters' attributes into variables for ego

proportion()

The share of alters with a given attribute: the share of advisers who are partners.

diversity()

Blau's index, 1 minus the sum of squared shares. 0 for one category, at most 0.5 for two and 0.67 for three.

mean()

The average of a continuous attribute among the alters, such as age or tenure.

Alters are the out-neighbors

In a directed network the alters of nwaltergen are the colleagues the lawyer turns to, because exposure follows the direction of the tie. The denominator of a share is the out-degree, not the size from nwego.

Shares and diversity as variables for ego, one office at a time

```
. nwaltergen pshare = proportion(alter.partner==1), net(lazega_adv)
. nwaltergen sh1 = proportion(alter.officenum==1), net(lazega_adv)
. nwaltergen sh2 = proportion(alter.officenum==2), net(lazega_adv)
. nwaltergen sh3 = proportion(alter.officenum==3), net(lazega_adv)
. generate sameoff = cond(officenum==1, sh1, cond(officenum==2, sh2, sh3))
. nwaltergen odiv = diversity(alter.officenum), net(lazega_adv)
```

One share per office, then pick ego's own

proportion() takes a fixed value. The share of alters in ego's own office is built from sh1, sh2 and sh3, and cond() chooses the one that fits ego's office.

Numeric groups

officenum (1 Boston, 2 Hartford, 3 Providence) and partner (1 for partners) are the numeric versions of office and status. Commands that take groups need them.

Everybody turns to partners, whatever their own rank

```
. tabstat pshare, by(partner) statistics(mean sd n)
associate |   .6039024   .2488415         35
  partner |    .662313   .1866759         36
```

Ego	advisers who are partners	partners among the 70 colleagues
associates	**60%**	51% (36 of 70)
partners	**66%**	50% (35 of 70)
all	**63%**	about half

Above the baseline in both groups. Both turn to partners more often than availability alone would predict, and the two groups differ only a little.

Office homophily is strong; mind the baseline and the diversity index

Office	Lawyers	Advisers from own office	Colleagues in own office	Mean diversity
Boston	48	**.879**	67% (47 of 70)	.190
Hartford	19	**.632**	26% (18 of 70)	.399
Providence	4	**.110**	4% (3 of 70)	.458

Every office draws on its own

Well beyond its share of the firm. Not all of it is preference: colleagues down the hall are available in a way colleagues in another city are not.

Diversity is not homophily

18 lawyers (16 Boston, 2 Hartford) have a diversity of zero. A Boston lawyer who asked only Hartford would score zero too. Use the own-office share for homophily.

Name the network: without net(), the current network is used

```
. nwaltergen pshare_wrong = proportion(alter.partner==1)
(no network name given: using the current network lazega_fr)
. summarize pshare pshare_wrong
    Variable |        Obs        Mean    Std. dev.       Min        Max
      pshare |         71     .633519    .2199212   .1111111          1
pshare_wrong |         69    .5753796    .2740652          0          1
```

69

lawyers instead of 71: the friendship network, not the advice network

Which network is current depends on what ran last

After nwwebuse and nwego it is the friendship network. After nwegoplot on lazega_adv it is lazega_adv, and the same omission gives the right answer by luck. Name the network with net(), and check one value by hand.

Check run of 17 Nov 2026; the chapter states the rule in words.

The same measures from a survey, with collapse

```
. clear
. input resp alter kin knows
   1 1 1 2
   1 2 1 2
   1 3 0 2
   1 4 0 0
   2 1 1 1
   2 2 0 1
   3 1 1 0
   end
```

Three respondents name four, two and one alters.

```
. collapse (count) size=alter (mean) kinshare=kin (sum) knows, by(resp)
. generate density = knows / (size * (size - 1))
(1 missing value generated)
. list, clean noobs
    resp   size   kinshare   knows   density
       1      4         .5       6         .5
       2      2         .5       2          1
       3      1          1       0          .
```

- **Each alter-alter tie is counted twice**

  So density = knows / (size x (size - 1)): three ties among the four alters of respondent 1 give .5, one tie between the two alters of respondent 2 gives 1. The single alter of respondent 3 leaves density missing, as nwego would.

Exposure: the share of alters with a behavior

```
. nwwebuse glasgow, nwclear
. generate byte smoker1 = (smoke1 > 1)
. nwaltergen frsmoke = proportion(alter.smoker1==1), net(glasgow1)
. nwaltergen frsmoke2 = proportion(alter.smoker1==1), net(glasgow1) hop(2)
. summarize frsmoke frsmoke2
    frsmoke |         46    .2507246    .3382408          0          1
   frsmoke2 |         38    .3412224    .3801773          0          1
```

46

of the 50 girls name a friend: only they have a share

25.1%

of a girl's friends smoke, on average

- **hop(2): exposure two steps out**

  The usual multi-hop question of diffusion research. Only 38 girls have anyone exactly two steps away. A smoker is a girl with occasional or regular smoking (smoke1 > 1).

Smokers have smoking friends, more so two steps out

```
. tabstat frsmoke frsmoke2, by(smoker1) ///
>         statistics(mean n)
```

Girl's own smoking	friends who smoke	girls two steps out who smoke
non-smokers	17% (34 girls)	24% (28 girls)
smokers	**47% (12 girls)**	**63% (10 girls)**

- **Reachable is not influential**

  That girls two steps away matter to ego needs a stated mechanism of transmission. The command in the chapter's tabstat line is shown on the left; the long line is continued over two lines in a do-file.

Exposure as a covariate: the timing comes first

```
. generate byte smoker2 = (smoke2 > 1)
. regress smoker2 frsmoke smoker1
   Number of obs   =        46
     smoker2 | Coefficient  Std. err.      t    P>|t|
     frsmoke |   .4692686   .1913345     2.45   0.018
     smoker1 |   .4034904   .1457721     2.77   0.008
       _cons |   .1249105   .0759978     1.64   0.108
```

.47

higher probability of smoking, all friends against none (standard error .19)

Linear probability model: coefficients are differences in probability.

- **Exposure comes before the outcome**

  Friends' smoking at wave 1, smoking at wave 2, the girl's own smoking at wave 1 held fixed. 46 girls: those who named a friend at wave 1. Confidence interval .08 to .86.

The coefficient is not an estimate of peer influence

It says

Girls whose friends smoked were more likely to smoke a year later than girls with the same earlier smoking whose friends did not, among 46 girls.

It does not say

That friends made them smoke. The girls chose their friends, and whatever led to the choice may also lead to smoking.

Selection and contagion

Homophily on unobserved traits and contagion are in general confounded in observational data.

Chapter 25

Models friendship choice and behavior change together: the honest way with three waves.

Five ways an ego-network analysis goes wrong

A missing measure shrinks the sample

Four girls name no friend: their exposure is missing and the regression drops them. Check the number of observations.

Four measures, one model

Size, density, effective size and constraint give unstable coefficients that cannot be read one at a time.

Composition read as preference

A circle of colleagues from one's own office is what the office makes available, as much as what ego chooses.

The wrong network

Without net(), nwaltergen takes the most recently created network. Name it, and check one value by hand.

diversity() on a continuous variable

Blau's index treats its argument as a few categories. On tenure or age use mean().

What to take away from the lecture

A unit of analysis

Cut from a whole network or collected one respondent at a time. Size, composition and density survive in a survey.

A baseline for composition

Read the circle against what the setting made available to ego.

Exposure comes before the outcome

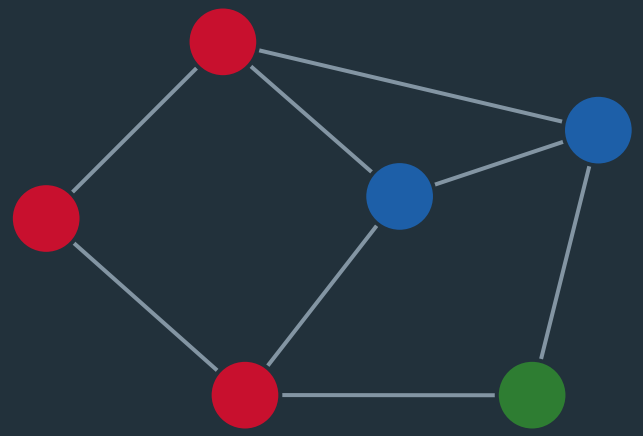
A covariate with the right causal direction, still confounded by the choice of contacts.

In the firm

Partners have larger circles (24.9 against 14.4), everybody turns to partners, and offices draw on their own lawyers.

Break

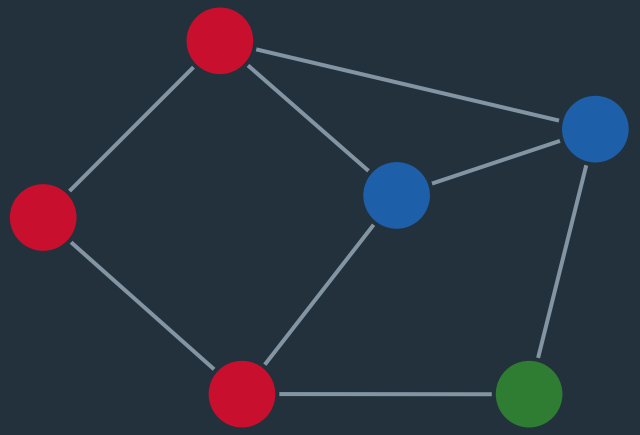
10 minutes



4

Lab: size, density and circles

45 min · Run along: nwego, correlate, tabstat, nwegoplot; Ch. 11, Ex. 1 and 4



Run along: size, density and two circles

```
. nwwebuse lazega, nwclear
. quietly nwdegree lazega_adv, generate(adv_out adv_in)
. nwego lazega_adv, sizevar(egosize) densvar(egodens)
. correlate egosize adv_out adv_in
. tabstat egosize egodens, by(partner) statistics(mean sd n)
. list _nwnode egosize egodens if inlist(_nwnode, "n62", "n70"), noobs
. nwegoplot lazega_adv, ego(n62) label(_nwnode) ///
>   color(partner) arrowfactor(0) scheme(s1network)
```

If a number differs

Check the network name in the command (lazega_adv, not lazega_fr or lazega_cow), and that nwclear came with nwwebuse. Try sortby(officenum) or hop(2) on the n70 plot when finished.

Checkpoints

egosize: mean 19.71831, range 6 to 44

egodens: mean .3153674

correlation with adv_out .9085, with adv_in .8083

partners 24.89 and .2966; associates 14.4 and .3346

n62: 12 alters, .530303; n70: 12 alters, .2575758

Hands-on (25 min)

Ch. 11, Ex. 1

An ego has four alters, and there are two ties among them in an undirected network. Compute the ego density, and then the effective size with the formula n - d(n-1). Check the result against Burt's form k - 2t/k from Chapter 8.

Hint Four alters form 4 x 3 / 2 = 6 pairs; the density is the share of these pairs that are tied. In Burt's form, k is the number of contacts and t the number of ties among them.

Ch. 11, Ex. 4

Burt's argument implies that a lawyer whose advisers are not tied to one another (low ego density) has low constraint. Check it in the law firm: compute ego density with nwego and constraint with nwburt on lazega_adv, and report their correlation. Then find the lawyer with the lowest ego density and report that lawyer's constraint, ego-network size and out-degree. Why can that lawyer's constraint still be high?

Hint nwego lazega_adv, sizevar(egosize) densvar(egodens), nwburt lazega_adv and nwdegree lazega_adv, generate(adv_out adv_in); then correlate egodens _constraint and gsort egodens. Recall which ties nwburt uses in a directed network (Week 6), and compare that with the alters that nwego counts.

Debrief Ex. 1: density 1/3, effective size 3, nearer 4 than 1

```
. clear
. input str3 a str3 b
   "E" "A"
   "E" "B"
   "E" "C"
   "E" "D"
   "A" "B"
   "C" "D"
   end
. nwset a b, edgelist undirected name(tiny)
```

```
. nwego tiny, sizevar(tsize) densvar(tdens)
. quietly nwburt tiny
. list _nwnode tsize tdens _effsize if _nwnode == "E", noobs
  | _nwnode   tsize      tdens   _effsize |
  |       E       4   .3333333          3 |
```

.33

ego density: 2 of 6 pairs

3

effective size: 4 - (1/3) x 3

Burt's form agrees

k - 2t/k = 4 - 2 x 2 / 4 = 3. Nearer 4 than 1: the alters are only moderately redundant. State undirected in nwset.

Debrief Ex. 4: density and constraint go together only loosely

.51

correlation of ego density with constraint, 71 lawyers

Lowest density	density	constraint	size	out-degree
n47	.143	.258	7	4
n22	.190	.081	30	21
n68	.199	.212	13	5
mean constraint		.171		

It says

Low density goes with low constraint on the whole (r = .51), but not for each lawyer: n47 has the lowest density and an above-average constraint.

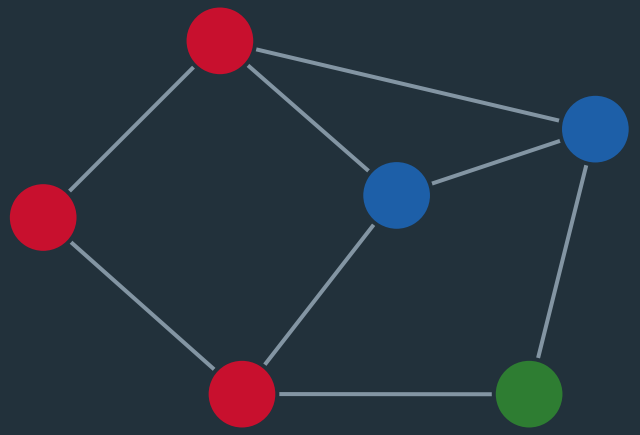
It does not say

That density alone settles constraint. nwburt uses the ties ego sends: n47 sends 4 while nwego counts 7 alters, and the few contacts take large shares of n47's relations.

5

Lab: composition, exposure and promotion

45 min · Run along: nwaltergen, own-office share, Glasgow exposure; Ch. 11, Ex. 6 and 9



Run along: composition by office

```
. nwaltergen pshare = proportion(alter.partner==1), net(lazega_adv)
. nwaltergen sh1 = proportion(alter.officenum==1), net(lazega_adv)
. nwaltergen sh2 = proportion(alter.officenum==2), net(lazega_adv)
. nwaltergen sh3 = proportion(alter.officenum==3), net(lazega_adv)
. generate sameoff = cond(officenum==1, sh1, cond(officenum==2, sh2, sh3))
. nwaltergen odiv = diversity(alter.officenum), net(lazega_adv)
. tabstat sameoff odiv, by(office) statistics(mean n)
. tabulate office if odiv == 0
```

If a number differs

Is net(lazega_adv) in every nwaltergen line? Is the cond() chain typed as printed?

Checkpoints

sameoff: Boston .8792543, Hartford .6320878, Providence .1095085

odiv: .1901006, .3986369, .4579772

18 lawyers with odiv = 0: 16 Boston, 2 Hartford

Run along: the exposure regression

```
. nwwebuse glasgow, nwclear
. generate byte smoker1 = (smoke1 > 1)
. nwaltergen frsmoke = proportion(alter.smoker1==1), net(glasgow1)
. generate byte smoker2 = (smoke2 > 1)
. regress smoker2 frsmoke smoker1
   Number of obs   =        46
     frsmoke |   .4692686   .1913345     2.45   0.018
     smoker1 |   .4034904   .1457721     2.77   0.008
```

Which girls are in?

Those who named at least one friend at wave 1. The regression drops the others without a word.

Checkpoints

frsmoke: 46 observations, mean .2507246

46 girls in the regression

frsmoke .4692686 (.1913345), p = .018

smoker1 .4034904 (.1457721)

Hands-on (25 min)

Ch. 11, Ex. 6

Rerun the exposure regression one wave later: regress smoking at the third wave on the share of smoking friends in the second-wave network glasgow2, holding the girl's second-wave smoking fixed. Report the exposure coefficient, its standard error and the number of girls, and compare them with the estimate for waves 1 to 2 (0.47, standard error 0.19, 46 girls). What would you need before calling either coefficient an effect of friends?

Hint Generate smoker2 = (smoke2 > 1) and smoker3 = (smoke3 > 1), then nwaltergen frsm_w2 = proportion(alter.smoker2==1), net(glasgow2) and regress smoker3 frsm_w2 smoker2.

Ch. 11, Ex. 9

A firm wants to know whether its junior staff with many contacts get promoted faster, and whether those contacts are well placed or merely numerous. Choose one of ego-network size, ego density, diversity() of the alters' departments and Burt's constraint as the main measure, and defend the choice in a paragraph. Name one of the others you would add as a check (and why only one), say when the network must be measured relative to the promotion, and state the result that would make you drop the hypothesis.

Hint Ask which of the four measures counts the contacts and which describes where they sit or how they are tied to one another. For the timing, ask whether a promotion could itself bring new contacts.

Debrief Ex. 6: a weaker estimate a wave later, the same caution

Smoking at the later wave on	waves 1 to 2	waves 2 to 3
exposure (share of smoking friends)	**.469**	**.287**
standard error	.191	.154
p	.018	.069
own smoking at the earlier wave	.403	.556
girls	46	47

With standard errors of .15 to .19 the two exposure estimates are not distinguishable from each other. Do not read the drop as a finding.

It says
A positive association between friends' smoking and a girl's own later smoking in both pairs of waves, weaker and less precise in the second.

It does not say
That friends make girls smoke. The girls chose their friends. Both estimates need a model of friendship choice and behavior change together (Chapter 25).

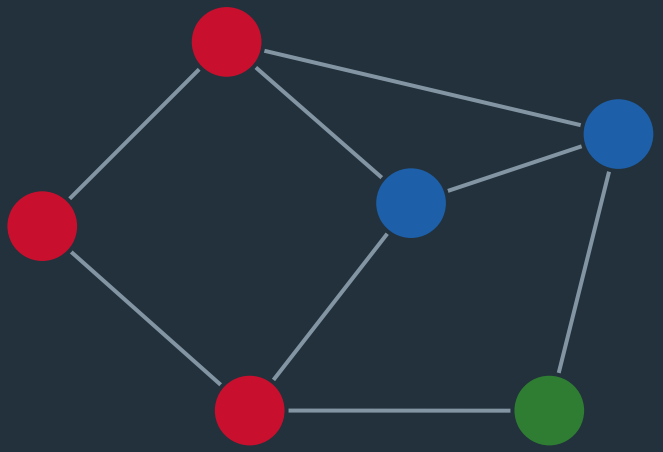
Debrief Ex. 9: choose the measure that fits the mechanism

Part of the answer	What a good paragraph contains
Main measure	one of size, density, diversity(), constraint, matched to the mechanism: numerous or well placed
One check	one other measure only (the four overlap), chosen so that it separates counting from position
Timing	the network measured before the promotion: a promotion can itself bring new contacts
Dropping the hypothesis	stated in advance: a coefficient near zero with the check in the model, or the opposite sign

A sketch, not the book's answer

The book gives hints for the open exercises, not solutions. Grade the reasoning: the mechanism, the overlap of the measures, the timing.

To read after the session



Before next week: read Chapter 15 (the sections on density and random graphs) and Chapter 18 (Random networks and null models)

Left over from Chapter 11: Ex. 2 and 3 (density and effective size by hand), Ex. 5 (law school), Ex. 7 and 8 (years with the firm, two-hop exposure), Ex. 10 (survey cap)

Start Stata, run nwwebuse lazega, nwclear and nwsummarize lazega_adv

Do-files and teaching material: nwcommands.org/book

Instructors are free to use, adapt and share these slides.